Novel compounds

The present invention relates to novel bacterial genes and processes for improving the manufacture of clavams e.g. clavulanic acid. The present invention also provides novel organisms capable of producing increased amounts of clavulanic acid.

Microorganisms, in particular Streptomyces *sp.* produce a number of antibiotics including clavulanic acid and other clavams, cephalosporins, polyketides, cephamycins, tunicamycin, holomycin and penicillins. There is considerable interest in being able to manipulate the absolute and relative amounts of these antibiotics produced by the microorganism and accordingly there have been a large number of studies investigating the metabolic and genetic mechanisms of the biosynthetic pathways [Domain, A.L. (1990) "Biosynthesis and regulation of beta-lactam antibiotics." In : 50 years of Penicillin applications, history and trends]. Many of the enzymes which carry out the various steps in the metabolic pathways and the genes which code for these enzymes are known.

Clavams can be arbitrarily divided into two groups dependent on their ring stereochemistry (5S and 5R clavams). The biochemical pathways for the biosynthesis of 5R and 5S clavams have not yet been fully elucidated but it has been suggested that they are derived from the same starter units (an as yet unidentified 3 carbon compound [Townsend,C.A. and Ho, M.F. (1985) J. Am. Chem. Soc. 107 (4), 1066-1068 and Elson, S.W. and Oliver, R.S. (1978) J. Antibiotics XXXI No.6, 568] and arginine [Valentine, B.P. *et al* (1993) J. Am Chem. Soc. 15, 1210-1211] and share some common intermediates [Iwata-Reuyl, D. and C.A.Townsend (1992) J.Am. Chem. Soc. 114: 2762-63, and Janc, J.W. *et al* (1993) Bioorg. Med. Chem. Lett. 3:2313-16].

Examples of 5S clavams include clavam-2-carboxylate (C2C), 2-hydroxymethylclavam (2HMC), 2-(3-alanyl)clavam, valclavam and clavaminic acid [GB 1585661 , Rohl, F. et al. Arch. Microbiol. 147:315-320, US 4,202,819] There are, however, few examples of 5R clavams and by far the most well known is the beta lactamase inhibitor clavulanic acid which is produced by the fermentation of *Streptomyces clavuligerus.* Clavulanic acid, in the form of potassium clavulanate is combined with the beta-lactam amoxycillin in the antibiotic AUGMENTIN (Trade

Mark SmithKline Beecham). Because of this commercial interest, investigations into the understanding of clavam biosynthesis have concentrated on the biosynthesis of the 5R clavam, clavulanic acid, by *S.clavuligerus*. A number of enzymes and their genes associated with the biosynthesis of clavulanic acid have been identified and published. Examples of such publications include Hodgson, J.E. *et al.*, Gene 166, 49-55 (1995), Aidoo, K.A. *et al.*, Gene 147, 41-46 (1994), Paradkar, A.S. *et al.*, J. Bact. 177(5), 1307-14 (1995). In contrast nothing is known about the biosynthesis and genetics of 5S clavams other than clavaminic acid which is a clavulanic acid precursor produced by the action of clavaminic acid synthase in the clavulanic acid biosynthetic pathway in *S. clavuligerus* .

Gene cloning experiments have identified that *S.clavuligerus* contains two clavaminic acid synthase isoenzymes, cas1 and cas2 [Marsh, E.N. *et al* Biochemistry 31, 12648-657, (1992)] both of which can contribute to clavulanic acid production under certain nutritional conditions [Paradkar, A.S. *et al.*, J. Bact. 177(5), 1307-14 (1995)]. Clavaminic acid synthase activity has also been detected in other clavulanic acid producing micro-organsims, ie. *S. jumonjinensis* [Vidal, C.M., ES 550549, (1987)] and *S .katsurahamanus* [Kitano, K. *et al.*, JP 53-104796, (1978)] as well as *S. antibioticos*, a producer of the 5S clavam, valclavam [Baldwin, J.E. *et al.*, Tetrahedron Letts. 35(17), 2783-86, (1994)] . The latter paper also reported *S. antibioticos* to have proclavaminic acid amidino hydrolase activity, another enzyme known to be involved in clavulanic acid biosynthesis. All other genes identified in *S.clavuligerus* as involved in clavam biosynthesis have been reported to be required for clavulanic acid biosynthesis [Hodgson, J.E. *et al.*, Gene 166, 49-55 (1995), Aidoo, K.A. *et al.*, Gene 147, 41-46 (1994)] and as yet none have been reported which are specific for the biosynthesis of 5S clavams.

We have now identified certain genes which are specific for the biosynthesis of 5S clavams as exemplified by C2C and 2HMC in *S. clavuligerus*. Accordingly the present invention provides DNA comprising one or more genes which are specific for 5S clavam biosynthesis in *S. clavuligerus* and which are not essential for 5R clavam (e.g. clavulanic acid) biosynthesis.

By "gene" as used herein we also include any regulatory region required for gene function or expression. In a preferred aspect the DNA is as identified as Figure

1. Preferably the DNA comprises the nucleotide sequences indicated in Figure 1 designated as orfup3, orfup2, orfup1, orfdwn1, orfdwn2 and orfdwn3. The present invention also provides proteins coded by said DNA. The present invention also provides vectors comprising the DNA of the invention and hosts containing such vectors.

Surprisingly we have found that when at least one of the genes according to the invention is defective the amount of clavulanic acid produced by the organism is increased. Accordingly the present invention also provides processes for increasing the amount of clavulanic acid produced by a suitable microorganism. In one aspect of the invention the genes identified can be manipulated to produce an organism capable of producing increased amounts of clavam, suitably clavulanic acid. The findings of the present work also allow an improved process for the identification of organisms with higher clavulanic acid production comprising a preliminary screening for organisms with low or no 5S clavam production (for example by hplc and/or clavam bioassay as described in the examples herein).

Suitably the 5*S* clavam genes of the present invention can be obtained by conventional cloning methods (such as PCR) based on the sequences provided herein. The function of the gene can be interfered with or eliminated/deleted by genetic techniques such as gene disruption [Aidoo, K.A. et al., (1994), Gene, 147, 41-46]., random mutagenesis, site directed mutagenesis and antisense RNA.

In a further aspect of the invention there are provided plasmids containing one or more defective genes, preferably the plasmids *pCEC060, pCEC061, pCEC056 and pCEC057,* described below.

Suitably, the plasmids of the invention are used to transform an organism such as *S. clavuligerus,* e.g. strain ATCC 27064 (which corresponds to *S. clavuligerus* NRRL 3585). Suitable transformation methods can be found in relevant sources including : Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989), *Molecular cloning: a laboratory manual, 2nd Ed.*, ColdSpring Harbor Laboratory, Cold Spring Harbor, N.Y ; Hopwood, D.A. *et al.* (1985), *Genetic Manipulation of Streptomyces. A Cloning Manual,* and Paradkar, A.S. and Jensen, S.E. (1995), J. Bacteriol. 177 (5): 1307-1314.

Strains of the species *S. clavuligerus* are used industrially to produce clavulanic acid (potassium clavulanate). Within the British and United States Pharmacopoeias for potassium clavulanate (British Pharmacopoeia 1993, Addendum 1994, p1362-3 and U.S. Pharmacopeia Official Monographs 1995, USP 23 NF18 p384-5) the amounts of the toxic 5S clavam, clavam-2-carboxylate, are specifically controlled.

Therefore in a further aspect of the invention there is provided an organism capable of producing high amounts of clavulanic acid but has been made unable to make C2C or capable of producing high amounts of clavulanic acid but able to make only low levels of C2C. Suitably the clavulanic acid producing organism contains one or more defective clavam genes, and is preferably the *S. clavuligerus* strain 56-1A, 56-3A, 57-2B, 57-1C, 60-1A, 60-2A, 60-3A, 61-1A, 61-2A, 61-3A, and 61-4A, described below. Such organisms are suitable for the production of clavulanic acid without the production of the 5S clavam, clavam-2-carboxylate or with significantly reduced production of clavam-2-carboxylate.

EXAMPLES

In the examples all methods are as in Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning A Laboratory Manual (2nd Edition), or Hopwood, D.A. et al. (1985) Genetic Manipulation of Streptomyces. A Cloning Manual, and Paradkar, A.S. and Jensen, S.E. (1995) J. Bacteriol. 177 (5): 1307-*1314* unless otherwise stated.

I. DNA sequencing of the *Streptomyces clavuligerus* chromosome upstream and downstream of the clavaminate synthase gene *cas1*.

A. Isolation of *cas1*.

To isolate chromosomal DNA fragments from *Streptomyces clavuligerus* NRRL 3585 encoding the gene for clavaminate synthase isozyme 1 (*cas1*) an oligonucleotide probe RMO1 was synthesised based on nucleotides 9-44 of the previously sequenced *cas1* gene (Marsh, E.N., Chang, M.D.T. and Townsend, C.A. (1992) Biochemistry 31: 12648-12657). Oligonucleotides were constructed using standard methods on an Applied Biosystems 391 DNA Synthesiser. The sequence of

RMO1, a 36-mer, was synthesised in the antiparallel sense to that published by Marsh et al (1992, ibid) RMO1 was radiolabelled with ^{32}P using standard techniques for end-labelling DNA oligonucleotides (Sambrook *et al.*, 1989 ibid), and was used to screen a cosmid bank of *Streptomyces clavuligerus* genomic DNA by Southern hybridization as described by Stahl and Amann (In : Nucleic acid techniques in bacterial systematics. Ed. E. Stackebrandt and M. Goodfellow. Toronto : John Wiley and Sons, p. 205-248, 1991). The genomic bank of *S. clavuligerus* DNA, prepared in cosmid pLAFR3, was as described by Doran, J.L *et al.*, (1990), J. Bacteriol. 172 (9), 4909-4918.

Colony blots of the *S. clavuligerus* cosmid bank were incubated overnight with radiolabelled RMO1 at 60°C in a solution consisting of 5 x SSC, 5 x Denhardt's solution, and 0.5% SDS (1 x SDS: 0.15 *M* NaCl + 0.015 *M* Na3citrate; 1 x Denhardt's solution: 0.02% BSA, 0.02% Ficoll, and 0.02% PVP). The blots were then washed at 68°C for 30 minutes in a solution of 0.5 x SSC + 0.1% SDS. One cosmid clone, 10D7, was isolated that hybridised strongly to RMO1 and gave hybridization signals upon digestion with restriction endonucleases *Sac*I and *Eco*RI that were consistent with hybridization signals detected in similar experiments with digests of *S. clavuligerus* genomic DNA.

B. DNA sequencing of the *S. clavuligerus* chromosome flanking *cas1*.

A partial restriction map of cosmid 10D7 was generated using restriction endonucleases *Sac*I, *Nco*I, and *Kpn*I. Southern hybridization experiments between RMO1 and various digests of 10D7 DNA indicated that *cas1* was most likely located at one end of a 7-kb *Sac*I-*Sac*I DNA subfragment. This fragment consisted of the *cas1* open reading frame and approximately 6 kb of upstream DNA. The 7-kb fragment was then subcloned from a *Sac*I digest of 10D7 in the phagemid vector pBluescriptII SK+ (2.96 kb; Stratagene), thus generating the recombinant plasmid pCEC007.

To facilitate the process of sequencing the chromosome upstream of *cas1*, a 3-kb *Nco*I-*Nco*I subfragment of the 7-kb *Sac*I-*Sac*I fragment was subcloned in pUC120 (3.2 kb; Vieirra and Messing, Methods Enzymol. 153, 3-11, 1987)) in both orientations, generating the recombinant plasmids pCEC026 and pCEC027. The 3-

kb subfragment consisted of the amino-terminal-encoding portion of *cas1* and approximately 2.6 kb of upstream DNA.

Nested, overlapping deletions were created in both pCEC026 and pCEC027 using exonuclease III and S1 nuclease digestion (Sambrook *et al.*, 1989 ibid) and the DNA sequence of the 3-kb *Nco*I-*Nco*I fragment was determined on both strands by the dideoxy chain termination method (Sanger, F., Nicklen, S. and Coulson, A.R. (1977), Proc. Natl. Acad. Sci. U.S.A. 74: 5463-5467) using a Taq dye-deoxy[a] terminator kit and an Applied Biosystems 373A Sequencer.

To determine the DNA sequence of the chromosome immediately downstream of *cas1* a 4.3-kb *Kpn*I-*Eco*RI DNA fragment was subcloned from cosmid clone 10D7 in pBluescriptII SK+, generating pCEC018. From pCEC018 a 3.7-kb *Sac*I-*Sac*I subfragment was cloned in pSL1180 (3.422 kb, Pharmacia); one of the *Sac*I termini of this fragment partially overlapped the TGA stop codon of *cas1*, the other was vector encoded. Both orientations of the 3.7-kb fragment were obtained during subcloning and the resulting recombinant plasmids were designated pCEC023 and pCEC024. Nested, overlapping deletions were created in both plasmids and the DNA sequence of the 3.7-kb fragment was determined on both strands. The nucleotide sequence of the *S. clavuligerus* chromosome generated in these experiments, including and flanking cas1 sequence is shown in Fig.1.

II. Functional analysis of the open reading frames flanking *cas1*.

Computer analysis of the DNA sequence upstream of *cas1* predicted the presence of two complete orfs and one incomplete orf . All three orfs were located on the opposite DNA strand to *cas1* and were thus oriented in the opposite direction. The first open reading frame, *orfup1*, was located 579 bp upstream of *cas1* and encoded a polypeptide of 344 amino acids (aa). The second open reading frame, *orfup2*, was located at 437 bp beyond the 3'-end of *orfup1* and encoded a 151 aa polypeptide. Beyond *orfup2* is *orfup3*. The start codon of *orfup3* overlaps the translational stop codon of *orfup2*, suggesting that the two orfs are translationally coupled. No translational stop codon for *orfup3* was located on the 3-kb *Nco*I-*Nco*I fragment.

A similar analysis of the DNA sequence downstream of *cas*1 predicted the presence of two complete orfs and one incomplete orf . Two of the orfs were located on the opposite DNA strand to *cas1* and were thus oriented towards *cas1*. The third orf was located on the same strand as *cas1* and was thus oriented away from it. The first downstream open reading frame, *orfdwn1*, was located 373 bp downstream of *cas1* and encoded a 328 aa polypeptide. The second open reading frame, *orfdwn2*, was located 55 bp upstream of *orfdwn1* and encoded a 394 aa polypeptide. At 315 bp upstream of *orfdwn2* and on the opposite strand was *orfdwn3*. Because no stop codon was observed for *orfdwn3* on the 3.7-kb fragment, it encoded an incomplete polypeptide of 219 aa.

Gene Disruption of the orfup and orfdwn open reading frames

To assess the possible roles of the open reading frames flanking *cas1* in the biosynthesis of clavulanic acid and the other clavams produced by *S. clavuligerus*, insertional inactivation or deletion mutants were created by gene replacement. The method used for gene disruption and replacement was essentially as described by Paradkar and Jensen (1995 ibid).

A. *orfup1*

A 1.5-kb *Nco*I-*Nco*I fragment carrying the apramycin resistance gene (*apr*^{r}), constructed as described in Paradkar and Jensen (1995 ibid), was treated with Klenow fragment to generate blunted termini (Sambrook *et al.*, 1989 ibid) and was ligated to pCEC026 that had been digested with *Bsa*BI and likewise treated with Klenow fragment. pCEC026 possesses a *Bsa*BI site located within *orfup1* at 636 bp from the translational start codon. The ligation mixture was used to transform competent cells of *E. coli* GM 2163 (available from New England Biolabs, USA., Marinus, M.G. *et al* M G G (1983) vol 122, p288-9) to apramycin resistance. From the resulting transformants two clones containing plasmids pCEC054 and pCEC055 were isolated; by restriction analysis pCEC054 was found to possess the *apr*^{r}-fragment inserted in the same orientation as *orfup1*, while pCEC055 possessed it in the opposite orientation.

To introduce pCEC054 into *S. clavuligerus*, plasmid DNA was digested with *Bam*HI and *Hin*dIII and ligated to the high-copy number *Streptomyces* vector pIJ486 (6.2 kb; Ward *et al.*, (1986) Mol. Gen. Genet. *203*: 468-478). The ligation mixture was then used to transform *E. coli* GM2163 competent cells to apramycin resistance. From the resulting transformants one clone, possessing the shuttle plasmid pCEC061, was isolated. This plasmid was then used to transform *S. clavuligerus NRRL 3585.* The resulting transformants were put through two successive rounds of sporulation on non-selective media and then replica plated to antibiotic containing media to identify apramycin-resistant and thiostrepton-sensitive transformants. From this process four putative mutants *(61-1A, -2A,-3A and -4A)* were chosen for further analysis.

To confirm that these putative mutants were disrupted in orfup1 genomic DNA was prepared from isolates 61-1A and 61-2A, digested with SacI and subjected to Southern blot analysis. The results of the Southern blot were consistent with a double cross-over having occurred and demonstrated that these mutants are true disruption replacement mutants in orfup1.

The mutants *61-1A, -2A, -3A and -4A* were grown in Soya-Flour medium and their culture supernatants were assayed by HPLC for clavulanic acid and clavam production. The composition of the Soya-Flour medium and the method for assaying clavams by HPLC were as previously reported (Paradkar and Jensen, 1995 ibid) except that the running buffer for the HPLC assay consisted of 0.1 *M* NaH_2PO_4 + 6% methanol, pH 3.68 (adjusted with glacial acetic acid). The HPLC analysis indicated that none of the mutants produced detectable levels of clavam-2-carboxylate or 2-hydroxymethylclavam. Furthermore, when culture supernatants were bioassayed against *Bacillus* sp. ATCC 27860, using the method of Pruess and Kellett (1983, J. Antibiot. 36: 208-212)., none of the mutants produced detectable levels of alanylclavam. In contrast, HPLC assays of the culture supernatants showed that the mutants appeared to produce superior levels of clavulanic acid when compared to the wild-type *(Table 1).*

Table 1

Clavulanic acid titre (CA) of orfup1 mutants in shake flask tests

STRAIN	*70 HOURS*	*70 HOURS*	*93 HOURS*	*93 HOURS*
	CA ug /ml	CA ug/mg DNA	CA ug /ml	CA ug/mg DNA
NRRL 3585 #1	87	915	166	1963
NRRL 3585 #2	66	790	159	1842
61-1A	272	2894	439	6113
61-2A	199	2148	225	2928
61-3A	54	692	221	2585
61-4A	0	0	226	2422

B. *orfdwn1* and *orfdwn2*

A deletion/replacement mutant in *orfdwn1* and *orfdwn2* was created by first digesting pCEC018 (7.3 kb) with *Nco*I and liberating a 1-kb subfragment containing most of *orfdwn1* and a portion of *orfdwn2* . The digest was fractionated by agarose-gel electrophoresis and the 6.3-kb fragment was excised and eluted from the gel. This fragment was then ligated to an *Nco*I-*Nco*I DNA fragment carrying *apr*^{r} and used to transform *E. coli* XL1-Blue to apramycin resistance. One clone was obtained from this experiment but restriction analysis of the resulting recombinant plasmid revealed that two copies of the apramycin resistance fragment had been ligated into the deletion plasmid. To eliminate the extra copy of the *apr*^{r}-fragment, the plasmid was digested with *Nco*I and self-ligated. The ligation mixture was used to transform *E. coli* GM2163 to apramycin resistance. From the transformants, two clones were isolated that contained plasmids pCEC052 and pCEC053 both of which possessed only one copy of the *apr*^{r}-fragment; pCEC052 possessed the *apr*^{r}-fragment inversely oriented with respect to *orfdwn1* and 2, while pCEC053 possessed the *apr*^{r}-fragment inserted in the same orientation as *orfdwn1* and 2.

A shuttle plasmid of pCEC052 was constructed by ligating *Bam*HI-digested pCEC052 with similarly digested pIJ486 and transforming *E. coli* GM2163 to apramycin resistance. From this experiment one clone was isolated that contained the shuttle plasmid pCEC060. This plasmid was used to transform wild-type *S. clavuligerus* 3585 to apramycin and thiostrepton resistance. The resulting transformants were put through two rounds of sporulation under non-selective

conditions and then replica plated to antibiotic containing media to identify apramycin resistant, thiostrepton sensitive colonies. Three putative mutants *(60-1A, -2A and -3A)* were chosen for further analysis.

To establish the identity of these putative mutants genomic DNA was isolated from strains 60-1A and 60-2A and digested with either SacI or BstEII and subjected to southern blot analysis.. The hybridisation bands generated from this experiment were consistent with both strains having undergone a double cross-over event demonstrating that these mutants are true disruption replacement mutants in orfdwn1/2.

When these were cultured in Soya-Flour medium and their culture supernatants assayed by HPLC, none of the mutants produced detectable levels of clavam-2-carboxylate or 2-hydroxymethylclavam. A bioassay of the culture supernatants showed that the mutants also failed to produce detectable levels of alanylclavam. As with the *orfup1* mutants, the *orfdwn1/2* mutants are capable of producing superior to wild-type levels of clavulanic acid *(Table2)*.

Table 2

Clavulanic acid titre (CA) of orfdwn1/2 mutants in shake flask tests

STRAIN	***70 HOURS***	***70 HOURS***	***93 HOURS***	***93 HOURS***
	CA ug/ml	*CA ug/mg DNA*	*CA ug/ml*	*CA ug/mg DNA*
NRRL 3585 #1	*87*	*915*	*166*	*1963*
NRRL 3585 #2	*66*	*790*	*159*	*1842*
60-1A	*164*	*1872*	*260*	*2911*
60-2A	*187*	*2013*	*108*	*1320*
60-3A	*79*	*994*	*214*	*2161*

orfdwn3

To disrupt orfdwn3 pCEC023 (consisting of a 3.7-kb fragment of cas1 downstream DNA subcloned into pSL1180) was digested with NcoI and then self ligated. After transforming *E.coli* with the ligation mixture a clone was isolated that possessed the plasmid pCEC031. This plasmid retained only the 1.9kb NcoI-EcoRI fragment encoding a portion of orfdwn2 and the incomplete orfdwn3. An

examination of the DNA sequence revealed that pCEC031 possessed a unique BstEII site at 158bp from the translational start site of orfdwn3. Therefore, pCEC031 was digested with BstEII, treated with Klenow fragment to create blunt ends and then ligated to a blunted apramycin resistance cassette. The ligation mixture was used to transform *E.coli* GM2163 to apramycin resistance and ampicillin resistance. Two transformants were selected that contained respectively pCEC050 and pCEC051. restriction analysis revealed that the apramycin resistance cassette was orientated in the same orientation as orfdwn3 in pCEC050 and in the opposite orientation in pCEC051. Both of these plasmids were then digested with HindIII and ligated to similarly digested pIJ486. The ligation mixtures were then used separately to transform *E.coli* GM2163 to apramycin and ampicillin resistance. The shuttle plasmids pCEC056 (pCEC050 + pIJ486) and pCEC057 (pCEC051+ pIJ486) were isolated from the resultant transformants. Both plasmids were then used to transform *S.clavuligerus* NRRL 3585.

One transformant was selected from each transformant experiment and put through two successive rounds of sporulation on non-selective media and then replica plated to antibiotic containing media to identify apramycin-resistant and thiostrepton-sensitive transformants. From this process two putative mutants were isolated from the progeny of each primary transformant. (56-1A and 56-3A for pCEC056, and 57-1C and 57-2B for pCEC057).

To establish the identity of these putative mutants genomic DNA was isolated from these strains and digested with either SacI or Acc65I and subjected to Southern blot analysis. The hybridisation bands generated from this experiment were consistent with both strains having undergone a double cross-over event demonstrating that these mutants are true disruption replacement mutants in orfdwn3.

When these strains were cultured in Soya-Flour medium and their culture supernatants assayed by HPLC, the mutants produced greatly reduced levels of clavam-2-carboxylate or 2-hydroxymethylclavam. A bioassay of the culture supernatants showed that the mutants also failed to produce detectable levels of alanylclavam. As with the orfup1 and orfdwn1/2 mutants, the orfdwn3 mutants were capable of producing superior to wild-type levels of clavulanic acid (Table 3).

Table 3

Clavulanic acid titre (CA) of orfdwn3 mutants in shake flask tests

STRAIN	***71 HOURS***	***71 HOURS***	***93 HOURS***	***93 HOURS***
	CA ug/ml	*CA ug/mg DNA*	*CA ug/ml*	*CA ug/mg DNA*
NRRL 3585 #1A	*180*	*1580*	*193*	*1790*
NRRL 3585 #1B	*179*	*1640*	*266*	*2310*
56-1A	*34*	*110*	*235*	*2160*
56-3A	*225*	*2140*	*274*	*2740*
57-1C	*253*	*2910*	*277*	*2920*
57-2B	*242*	*2240*	*193*	*1860*

The application discloses the following nucleotide sequences:

SEQ ID No. 1 : DNA sequence of Figure 1

SEQ ID No. 2 : orfup3 sequence

SEQ ID No. 3 : orfup2 sequence

SEQ ID No. 4 : orfup1 sequence

SEQ ID No. 5 : orfdwn1 sequence

SEQ ID No. 6 : ofrdwn2 sequence

SEQ ID No. 7 : orfdwn3 sequence

Figure 1: Nucleotide sequence of the *S. clavuligerus* chromosome including and flanking *cas1*

NcoI

```
1
GGTACCGCCCGCCGCCGACGGGGCCTCGGAGCCGGCCTGGCCACTGGTCCTGGTGGGGCC   60
          M  A  P  P  P  Q  G  P  A  E  A  P  G  T  V  L  V  V  G
61
ACCCTATCACCGGGCGGTGGGCCGCGTCGTCTGAGGGCCTGTGCCTGGGCACCCACACGC  120
       T  P  Y  H  G  A  V  R  R  L  L  S  G  S  V  S  G  H  T  H
```

<orfup3

```
    121
GCCTTTCCGGGCCTCCGGCCCAGTGTCGGTGCCCATTGCGCGCCACAGGAACGGGCGCAT   180
                              *  L  W  P  Y  R  A  T  D  K  G  A
Y
5          A  S  L  G  P  P  R  T  M
                   .        .        .        .        .
.
    181
TAGCCCCAGGTCTATCTGCTTCCGGGCCACCTGCTCCTTCAGGGCGTGGAGCATCTGGCA   240
10            D  P  D  L  Y  V  F  A  R  H  V  L  F  D  R  V  E  Y  V
T
                   .        .        .        .        .
.
    241
15  CGTGGTCGCGGGCCGCCGGGTGAGCCCCAGTGGGCGGGCGGTGCCGGGCAGGGCCACGAG   300
              C  W  R  G  A  A  W  E  P  D  G  A  R  W  P  G  D  R  H
E
                   .        .        .        .        .
.
20      301
TGGCACCCACCACGGGAGGCGCCGCTCCTCAAGCCAGGGCCAGTCTTAGGTCAACTGCCT   360
              G  H  T  T  G  E  A  A  L  L  E  T  G  T  L  I  W  N  V
S
                   .        .        .        .        .
25  .
    361
GGTGTCTACCACCCACTAGCTCGCCTACCACGGGGGCTCCAGCAGCTTCTCGGCCCGCTA   420
              W  L  H  H  T  I  S  R  I  T  G  G  L  D  D  F  L  R  A
I
30                 .        .        .        .        .
.
    421
GAGCCTGAACGGGGCCCGGTCTGGGGTGAACCCCTTCTTCTTCTGGCGCAGGAGCCGCTT   480
              E  S  K  G  R  A  L  G  W  K  P  F  F  F  V  A  D  E  A
35  F
                   .        .        .        .        .
.
    481
CATCAGCTAGCGCCCCCACGGCAGCGACGGCTGCGGCGGCAACAGCTTGCGGAACTTCAT   540
40            Y  D  I  A  P  T  G  D  S  G  V  G  G  N  D  F  A  K  F
Y

                   .        .        .        .        .
<orfup2
```

```
    541
GCGCCACTACTGGCGGAACGCGACGAGCAGGCAGTATGGCCGGCTACGGTGCCTGTACTT   600
          A  T  I  V  A  K  R  Q  E  D  T  M  G  A  S  A  V  S  M
                .        .        .        .        .
.
    601
TGCTGGAGGTCTCTAAGGCCCACCGACACGACCCCGACGCCTTCCCCACAGGGGGCGCTT   660
                .        .        .        .        .
.
    661
CCTGCCGCCTGCGGCGCCTGCGGCGCCGGCAGAGGGGCCGCCTGCCCAGGGTCGCAGGAC   720
                .        .        .        .        .
.
    721
CTCTCCCGAACCGCCGCCGAACTGCGGCACGACAGGGCGCCGAACGCCTTGCGCTTCATG   780
                .        .        .        .        .
.
    781
GCCGGTCGCATGCCCGCAACGTGGCCTGCACATGCGGCCAGCCCTGGGGAGCATGGGGGC   840
                .        .        .        .        .
.
    841
CTCGGCCGGCTGGGGCCGCCGAGGCCCCCATGCCTGCGCGGCCTGGCCGGGCTCGCTCGG   900
                .        .        .        .        .
.
    901
CCTGCCCAGCCTGCCACGCGCACCAAGGCCACACAGCCTGTCGAGCCTGCCTGGCCTGCC   960
                .        .        .        .        .
.
    961
ACGCGCACCAAGGCCACACAGCCTGTCGAGCCTGCCCAGCCTGCCACGCGCACCAAGGCC  1020
                .        .        .        .        .
.
   1021
GTGCGGCCTGCCCAGTCAACGGCTAGTACCGCTCGTTACGGCCCCACATGGCGAGGGGCC  1080
                      *  N  G  I  M  A  L  L  A  P  T  Y  R  E  G
                .        .        .        .        .
.
   1081
TGTGGCCCACCCTCTAGCGCCGGCAGTGGAGGCGCTCCCTGGCCAGCAGGTCGGCCTAGC  1140
          S  V  P  H  S  I  A  A  T  V  E  A  L  S  R  D  D  L  R  I
                .        .        .        .        .
.
   1141
TCCGCCGCCGCTCTAACAGGCGCTCTACCCGGCCCAAGCGCCACGGGCCCTAGCCCTGCT  1200
```

```
          S  A  A  A  L  N  D  A  L  H  A  P  N  A  T  G  P  I  P  V
                .        .        .        .        .
.
   1201
GCAGGAGCGGGGGCCACCACGTCGGTCCGCTCGCGCTCGACACGGTCCCAGTCGGGGTCTG   1260
          V  D  E  G  R  H  H  L  W  A  L  A  L  Q  A  L  T  L  G  L
                .        .        .        .        .
.
   1261
GCAGGCGCTGGCCCGCGTCGGCCACGTCGTTGCTCGCCAACGCGCGCTCCCGGCCTCGCG   1320
          G  D  A  V  P  R  L  R  H  L  L  S  R  N  R  A  L  A  P  A
                .        .        .        .        .
.
   1321
ACTTGGCCCCGACCGGGGCCGCCTTCAGGAGCAGGGGGTCTAGCAGCCACCACGCCTACC   1380
          S  F  R  P  Q  G  R  R  F  D  E  D  G  L  D  D  T  T  R  I
                .        .        .        .        .
.
   1381
ACGGCCACTCTTTTGGGGCAGGGTCTCCCCCGCATTCGCTGCTAGGGCTAGGGGTCGAGGG   1440
          T  G  T  L  F  G  R  G  L  P  A  Y  A  V  I  G  I  G  L  E
                .        .        .        .        .
.
   1441
CCGTCTGCCCGTGGTGGAGCAGGAGCTAGGGCGCGCTGGTGTCCGAGGTGAGCGAGACGT   1500
          R  C  V  P  V  V  E  D  E  I  G  R  S  W  L  S  W  E  S  Q
                .        .        .        .        .
.
   1501
GGCGGCAGTGGCCCACGTGGCGCAGGCGGGCCGCGTCGCACCGGCGCCTCCCGAGCCTCT   1560
          V  A  T  V  P  H  V  A  D  A  R  R  L  T  A  A  S  P  E  S
                .        .        .        .        .
.
   1561
CTGGCTCGGACGCCTGGAACGGGAGCGCGTGGTCGAGCCGGTGGCGTGGGTGCCAGAGGA   1620
          L  G  L  R  R  V  K  G  E  R  V  L  E  A  V  A  G  V  T  E
                .        .        .        .        .
.
   1621
GCTAGCCGTGGCGGCCCAGGCAGGTCACGACCATCATGTCCAGCTACGCCAGCCACGGCT   1680
          E  I  P  V  A  P  D  T  W  H  Q  Y  Y  L  D  I  R  D  T  G
                .        .        .        .        .
.
   1681
CTGCTGCGTCCCTGGCAAGCGTCCGGCGCGCCTGCATCCTGCCGAGCGGCGTGTTCGGGA   1740
```

```
          L  R  R  L  S  R  E  C  A  A  R  V  Y  S  P  E  G  C  L  G
                   .        .        .        .        .

      1741
CCCTCCGCGGCAGCCTGCTCGCGTGGTACGGCTTGAACCACCGCTAGTCGTGGAGCAGGG   1800
          Q  S  A  G  D  S  S  R  V  M  G  F  K  T  A  I  L  V  E  D
                   .        .        .        .        .

      1801
CCGCCGGGCGCTGGCGGGCAGGCTCGTCGAGGAGTGGCCGCGGCTCGGGGACCTGCAGCC   1860
          R  R  G  A  V  A  R  G  L  L  E  E  G  A  G  L  G  Q  V  D
                   .        .        .        .        .

      1861
GCCACAGGTCGTCCCACTGGGGCCGCAGCTGCCGCCGCGCCTACCACCGGCAGCGGGCCC   1920
          A  T  D  L  L  T  V  G  A  D  V  A  A  R  I  T  A  T  A  R
                   .        .        .        .        .

      1921
GCGCCAGGCCCGCAGGCATCTTCAGCCACCAGCCGTCCGTCGGCTCGGGGACCCGTGACT   1980
          A  R  D  P  R  G  Y  F  D  T  T  P  L  C  G  L  G  Q  A  S
                   .        .        .        .        .

      1981
GGCCTTCCAGGGCGTCCCGCGCCTGGCCGCCTGCGCCTTGGCGCCGCCTGTGCCTTGGCC   2040
          V  P  L  D  R  L  A  R  V  P  P  R  P  V  A  A  S  V  S  G
                   .        .     <orfup1     .        .

      2041
CCGGGGACTCGGGCGGAGAGCGGGACATACGGAACCTCCACAGGCGGAGCCGGGAACGGG   2100

GGCCCCTGAGCCCGCCTCTCGCCCTGTATGCCTTGGAGGTGTCCGCCTCGGCCCTTGCCC
          A  P  S  E  P  P  S  R  S  M
                   .        .        .        .        .

      2101
ACGAGGGCGAGGACGGGACGGAACGAAGGAGAGGACGGGACGGACAGCACGGACGGGACG   2160

TGCTCCCGCTCCTGCCCTGCCTTGCTTCCTCTCCTGCCCTGCCTGTCGTGCCTGCCCTGC

                   .        .        .        .        .

      2161
GACGGAACGGAGTCGGGAACCGGGGGGGGTGACCGGAACCGGGCCGTCCTTGGCCCTCCC   2220
```

```
CTGCCTTGCCTCAGCCCTTGGCCCCCCCCACTGGCCTTGGCCCGGCAGGAACCGGGAGGG
                    .         .         .         .         .
.
   2221
CCGTCCTCCCCGCCATCCGCCGTTCTCCCCCGTTCCCTCTCCCGTCCTCCAGCCAACACC   2280

GGCAGGAGGGGCGGTAGGCGGCAAGAGGGGGCAAGGGAGAGGGCAGGAGGTCGGTTGTGG
                    .         .         .         .         .
.
   2281
GCCGCCCTTTCCAAGCGCTTGACACGGCACCGACAGCCGCCGCCGGGCGCCCGATGGGGA   2340

CGGCGGGAAAGGTTCGCGAACTGTGCCGTGGCTGTCGGCGGCGGCCCGCGGGCTACCCCT
                    .         .         .         .         .
.
   2341
CCCGTGCCCGCCGGTGAGCGGCGGTGAGCGCCGGTACGGGACCCCACGCGCCGCCGCCCG   2400

GGGCACGGGCGGCCACTCGCCGCCACTCGCGGCCATGCCCTGGGGTGCGCGGCGGCGGGC
                    .         .         .         .         .
.
   2401
GGCGCCCGCCAGGGCCCGCGCGGCCACCCCGGCCCGCCCCGGCCGGAGCGGCGATCCGGG   2460

CCGCGGGCGGTCCCGGGCGCGCCGGTGGGGCCGGGCGGGGCCGGCCTCGCCGCTAGGCCC
                    .         .         .         .         .
.
   2461
CCGCTCGCTGCAAGAGGAACATCCACAGCCGCACAAGGAGCGCTCCGCACAGTGGGCACC   2520

GGCGAGCGACGTTCTCCTTGTAGGTGTCGGCGTGTTCCTCGCGAGGCGTGTCACCCGTGG
                    .         .         .         .         .
.
   2521
ACGTCCGCCCCGTCCCCCACACCGTGGCCGGTCCCCACCGGACAGCACAGCACCGCACAG   2580

TGCAGGCGGGGCAGGGGGTGTGGCACCGGCCAGGGGTGGCCTGTCGTGTCGTGGCGTGTC
                    .         .         .         .         .
.
   2581
CACCACATCGCACGGCACAGCACAGCACCACCGGCACGAGGAACCAAGGAAAGGAACCAC   2640

GTGGTGTAGCGTGCCGTGTCGTGTCGTGGTGGCCGTGCTCCTTGGTTCCTTTCCTTGGTG
```

cas1>

M T S V D C T A Y G P E L R A L A A

2641

ACCACCATGACCTCAGTGGACTGCACCGCGTACGGCCCCGAGCTGCGCGCGCTCGCCGCC. 2700

TGGTGGTACTGGAGTCACCTGACGTGGCGCATGCCGGGGCTCGACGCGCGCGAGCGGCGG

. . . . .

2701

CGGCTGCCCCGGACCCCCCGGGCCGACCTGTACGCCTTCCTGGACGCCGCGCACACAGCC 2760

R L P R T P R A D L Y A F L D A A H T A

. . . . .

2761

GCCGCCTCGCTCCCCGGCGCCCTCGCCACCGCGCTGGACACCTTCAACGCCGAGGGCAGC 2820

A A S L P G A L A T A L D T F N A E G S

. . . . .

2821

GAGGACGGCCATCTGCTGCTGCGCGGCCTCCCGGTGGAGGCCGACGCCGACCTCCCCACC 2880

E D G H L L L R G L P V E A D A D L P T

. . . . *Nco*I .

2881

ACCCCGAGCAGCACCCCGGCGCCCGAGGACCGCTCCCTGCTGACCATGGAGGCCATGCTC 2940

T P S S T P A P E D R S L L T M E A M L

. . . *Kpn*I. .

2941

GGACTGGTGGGCCGCCGGCTCGGTCTGCACACGGGGTACCGGGAGCTGCGCTCGGGCACG 3000

G L V G R R L G L H T G Y R E L R S G T

. . . . .

3001

GTCTACCACGACGTGTACCCGTCGCCCGGCGCGCACCACCTGTCCTCGGAGACCTCCGAG 3060

V Y H D V Y P S P G A H H L S S E T S E

. . . . .

3061

ACGCTGCTGGAGTTCCACACGGAGATGGCCTACCACCGGCTCCAGCCGAACTACGTCATG 3120

T L L E F H T E M A Y H R L Q P N Y V M

. . . . .

```
   3121
CTGGCCTGCTCCCGGGCCGACCACGAGCGCACGGCGGCCACACTCGTCGCCTCGGTCCGC   3180
        L  A  C  S  R  A  D  H  E  R  T  A  A  T  L  V  A  S  V  R
                 .        .        .        .        .
.
   3181
AAGGCGCTGCCCCTGCTGGACGAGAGGACCCGGGCCCGGCTCCTCGACCGGAGGATGCCC   3240
        K  A  L  P  L  L  D  E  R  T  R  A  R  L  L  D  R  R  M  P
                 .        .        .        .        .
.
   3241
TGCTGCGTGGATGTGGCCTTCCGCGGCGGGGTGGACGACCCGGGCGCCATCGCCCAGGTC   3300
        C  C  V  D  V  A  F  R  G  G  V  D  D  P  G  A  I  A  Q  V
                 .        .        .        .        .
.
   3301
AAACCGCTCTACGGGGACGCGGACGATCCCTTCCTCGGGTACGACCGCGAGCTGCTGGCG   3360
        K  P  L  Y  G  D  A  D  D  P  F  L  G  Y  D  R  E  L  L  A
                 .        .        .        .        .
.
   3361
CCGGAGGACCCCGCGGACAAGGAGGCCGTCGCCGCCCTGTCCAAGGCGCTCGACGAGGTC   3420
        P  E  D  P  A  D  K  E  A  V  A  A  L  S  K  A  L  D  E  V
                 .        .        .        .        .
.
   3421
ACGGAGGCGGTGTATCTGGAGCCCGGCGATCTGCTGATCGTCGACAACTTCCGCACCACG   3480
        T  E  A  V  Y  L  E  P  G  D  L  L  I  V  D  N  F  R  T  T
                 .        .        .        .        .
.
   3481
CACGCGCGGACGCCGTTCTCGCCCCGCTGGGACGGGAAGGACCGCTGGCTGCACCGCGTC   3540
        H  A  R  T  P  F  S  P  R  W  D  G  K  D  R  W  L  H  R  V
                 .        .        .        .        .
.
   3541
TACATCCGCACCGACCGCAATGGACAGCTCTCCGGCGGCGAGCGCGCGGGCGACGTCGTC   3600
        Y  I  R  T  D  R  N  G  Q  L  S  G  G  E  R  A  G  D  V  V

        A  F  T  P  R  G  *  SacI        .        .        .
.
   3601
GCCTTCACACCGCGCGGCTGAGCTCCCGGGTCCGACACCGCGCGGCTGAACCCACGGTCC   3660

CGGAAGTGTGGCGCGCCGACTCGAGGGCCCAGGCTGTGGCGCGCCGACTTGGGTGCCAGG
```

```
                    .         .         .         .         .
   .
      3661
   GGGGCCCACGGTCCGGCACCGCGCGGCTGAGCCCCCGGGTCCGGCAGCGGGCGGCTGAAC   3720
5
   CCCCGGGTGCCAGGCCGTGGCGCGCCGACTCGGGGGCCCAGGCCGTCGCCCGCCGACTTG
                    .         .         .         .         .
   .
      3721
10 CCCCGCCCCGGGCCACCGCCCGACCGCCCCCGCGCACCGGACGCGCCCGCCTGTACGGCG   3780

   GGGGCGGGGCCCGGTGGCGGGCTGGCGGGGGCGCGTGGCCTGCGCGGGCGGACATGCCGC
                    .         .         .         .         .
   .
15    3781
   GTCCCGCCCGGGCCCGTACACCTGAAGCGCCCGGCGGACCGCCGCCCCGCCGGGGGACGG   3840

   CAGGGCGGGCCCGGGCATGTGGACTTCGCGGGCCGCCTGGCGGCGGGGCGGCCCCCTGCC

20            ----------------->    <-----------------        .
   .
      3841
   ACAGAGCCGGGTGCGGGAGGACGTCCTCCCGCACCCGGCTCCCACCGTTCCGCACCGACC   3900

25 TGTCTCGGCCCACGCCCTCCTGCAGGAGGGCGTGGGCCGAGGGTGGCAAGGCGTGGCTGG
                    .         .         .         .         .
   .
      3901
   GCACCCGACCGTGCCGCAGGCGCCACCGGCACCGCACCGCCCGCGCCGGCAGCCACCACA   3960
30
   CGTGGGCTGGCACGGCGTCCGCGGTGGCCGTGGCGTGGCGGGCGCGGCCGTCGGTGGTGT
                    .         .         .         .         .
   .
      3961
35 GGCGCCACGCCGCCCGCACGGTGCCCGCGCTGCTCAGCCCCCGTCCACCGGGCTGTCCAG   4020

   CCGCGGTGCGGCGGGCGTGCCACGGGCGCGACGAGTCGGGGGCAGGTGGCCCGACAGGTC
                                          *  G  G  D  V  P  S  D
   L
40                  .         .         .         .         .
   .
      4021
   GTCGGCGGCGTCGCGCGGGGGCTACTTGAGGGCCAGCCGCCGGCTGGGGGGCCTGGGGCG   4080
            L  R  R  L  A  G  G  I  F  E  R  D  A  A  S  G  G  S  G
45 A
```

```
                    .         .         .         .         .
.
   4081
CTCTACGGGGGTGTGAGGGCCCTAGTGGAGGTCGCTCCGTATGCCGTCGTCTAGCCGGTG   4140
          L  H  G  W  V  G  P  I  V  E  L  S  A  Y  P  L  L  D  A
V
                    .         .         .         .         .
.
   4141
GGCGAAGAGCAGGAGCTGCCGCTTTGTGTGCAGGTCCCGCGGGCCGTCGTGGTGCCGGGC   4200
          R  K  E  D  E  V  A  F  C  V  D  L  A  G  P  L  V  V  A
R
                    .         .         .         .         .
.
   4201
GCGGCACTGCCTCCGGTCGCGGCGGAGCTGCGAGGGGGGCCGGGGCCCACAGCGGGGGTG   4260
          A  T  V  S  A  L  A  A  E  V  S  G  G  A  G  P  T  A  G
V
                    .         .   NcoI  .         .         .
.
   4261
TAGGCACAAGAGGGTCCACGCGTGGTACCACTCGTCTAGGCGCCGCGGCCCGGGCCTCTC   4320
          D  T  N  E  W  T  R  V  M  T  L  L  D  A  A  G  P  G  S
L
                    .         .         .         .         .
.
   4321
CTTCTGGACGAGGGTCTTCGGCCACTCCATGAGGAGCGCCCACCGCTTTGGGTCGAGGGC   4380
          F  V  Q  E  W  F  G  T  L  Y  E  E  R  T  A  F  G  L  E
R
                    .         .         .         .         .
.
   4381
CACCCGTGCCGCCCGGGTCTTCCTTGCGCTCCAGGGGGTGGGCCGCTTGTGGGCCGGGCG   4440
          H  A  R  R  A  W  F  S  R  S  T  G  W  G  A  F  V  R  G
A
                    .         .         .         .         .
.
   4441
GCGGAAGGCGGGGGCGAGGGGCCGCAGCCGCGACTCGCGGCGCCGGTCTGGCCTGTCGTC   4500
            A  K  R  G  R  E  G  A  D  A  S  L  A  A  A  L  G  S  L
L
                    .         .         .         .         .
.
```

```
   4501
CTGGTCCGACACGCCCGACGAGTGGCCGCGGGGCGTCTAGCCCCGCTAGGCCGCGTGGTA   4560
          V  L  S  H  P  S  S  V  P  A  G  C  I  P  A  I  R  R  V
M
                .        .        .        .        .
.
   4561
GGGGCCTACGCTGTGCCGGGTGACCATCCGCACCCGGCGCGGGTAGCTGGTCGGGCACTG   4620
          G  P  H  S  V  A  W  Q  Y  A  H  A  A  G  M  S  W  G  T
V
                .        .        .        .        .
.
   4621
GTCCCGGTCAAGGGCATGGGGGTCGAGGAGCCACTCGTCGGCCACGACGCGGCGCTGTAA   4680
          L  A  L  E  R  V  G  L  E  E  T  L  L  R  H  Q  A  A  V
N
                .        .        .        .        .
.
   4681
CAGGACGCCTCACTAGTCGCCTTTCGCCCTGGGGCTGCCCACCAACGGCCCGCTCGACCT   4740
          D  Q  P  T  I  L  P  F  R  S  G  S  P  H  N  G  P  S  S
S
                .        .        .        .        .
.
   4741
CTGGGGCAACGGCTTCTCAGGCCGCCACTGCTGCGTCATGGCGGCCCACAGGTCGCCGTC   4800
          V  G  N  G  F  L  G  A  T  V  V  C  Y  R  R  T  D  L  P
L
                .        .        .        .        .
.
   4801
GGGGCGTGGCTAGTCGGTCAGCATGGGCCACACCAGGGCCGGCTTCTTGCTGCCTGTCTC   4860
          G  A  G  I  L  W  D  Y  G  T  H  D  R  G  F  F  S  P  C
L
                .        .        .        .        .
.
   4861
GTGGTGCAAGCAGGGCAGCCGCAAGCCGCACGGCATGTACCGCATTGGCTAGGCCCGCAG   4920
          V  V  N  T  G  D  A  N  P  T  G  Y  M  A  Y  G  I  R  A
D

                .        .        .        .        .
<orfdwn1
   4921
GGCGTCCTGGAGGGGCAGGTCGTTGCCGTCAAGCAGCTAGAGCTTATACGCCGTAAGGTG   4980
```

```
                R  L  V  E  G  D  L  L  P  L  E  D  I  E  F  I  R  C  E
M
                         .        .        .        .        .
.
    4981
GCGACTGGAGGAACAAGCTAGGGGGGCCTGTTGTCCAGCCAGCACCGGCCTCTGAGTCTC   5040
                                                                  *
L
                         .        .        .        .        .
.
    5041
GGTCAACCCCCGCTAGAGCCACCGGGTGTCGAGGTCCGACGCGTCGACCTGTAGCACGCC   5100
                W  N  P  A  I  E  T  A  W  L  E  L  S  R  L  Q  V  D  H
P
                         .        .        .        .        .
.
    5101
CTAGTCGGGCCTCATGACCGTGACCTCGTCTATGAGGCCTAGCACGGCGAGGTGGTCGAA   5160
                I  L  G  S  Y  Q  C  Q  L  L  Y  E  P  D  H  R  E  V  L
K
                         .        .        .        .        .
.
    5161
GAGCTAGTACGCCAACTACAGCAGGCCCCACGGCTGGGTGAGGTCGGGGGCCAGCTGGTC   5220
                E  I  M  R  N  I  D  D  P  T  G  V  W  E  L  G  R  D  V
L
                         .        .        .        .        .
.
    5221
CCAGAACATCAGGCTCGGCTAGCCTGGGCAGAGCGGCCAGCGCGCGTCGCGGAGCCACTT   5280
                T  K  Y  D  S  G  I  P  G  T  E  G  T  A  R  L  A  E  T
F
                 NcoI    .        .        .        .        .
.
    5281
CGGGTACCCCGGCTTGGTCAAGAGCTTCTACTTCGGCGGCGGCGCCCTGCGGGTCACCAC   5340
                G  M  P  G  F  W  N  E  F  I  F  G  G  G  R  S  A  W  H
H
                         .        .        .        .        .
.
    5341
CCGGAGCGGCCTCAGGGCCCTCTGGTCCTGCAGGAAGTAGTGGGGCTGGGCGAGCGGGGC   5400
                A  E  G  S  D  R  S  V  L  V  D  K  M  V  G  V  R  E  G
R
```

. . . . .

.

5401
GGCGTCCCACGGCACCGGGCGGCGGAGCCGGAGGAGGGCCATCTACAGGTAGTCGGCCCG 5460
R L T G H G A A E A E E R Y I D M L R
A

. . . . .

.

5461
CTGCTAGACCAGCAGCCACAAGTAGTCCTAGCCGTGGTGCGGAGGGCCCGTGTCTTGGC 5520
V I Q D D T N M L I P V V G E R A C F
R

. . . . .

.

5521
CTTGCACAGGAGTGACTTCGACTTGCCGACCTTCTGCCCGCCCACCCCCGCGACCATCCC 5580
F T D E S F S F P Q F V P P H P R Q Y
P

. . . . .

.

5581
GAACCCGCGCTACGGGTGGAGCGCCTACTGCGGCAAGAGCAGCTCCGGGGCCGGCATCGC 5640
K P A I G V E R I V G N E D L G R G Y
R

. . . . .

.

5641
CGCGTGGCGGAGCATCCCCTTGAGGTCCAGGCCGTGGCCCTAGCAGGTGACGAGGGGCCT 5700
R V A E Y P F E L D P V P I T W Q E G
S

. . . . .

.

5701
CACCCACTTGCAGAGCCAGCAGGTGCGGAAGAACTACTAGAGGGTCACGAGGAGCTTCTC 5760
H T F T E T T W A K K I I E W H E E F
L

. . . . .

.

5761
CCGTGCTAACGCGGCCAGGGCGAGGGGCCGCAGCCTGTCCCACGGCGGCTGGGGCATGTG 5820
A R N R R D R E G A D S L T G G V G Y
V

. . . . .

.

```
   5821
GACGGGGTACTACAGCCGGGTCGCGAAGACCTTGGGCGCGCGCTAGGGCTGCTTCCGCGC   5880
          Q  G  M  I  D  A  W  R  K  Q  F  G  R  A  I  G  V  F  A
R
                .        .        .        .        .
.
   5881
CGGGGCCCAGTACACCAGCTCGTAGCGGTCTAGGAGCCGGTCGGCGTCGCCTAACACGTC   5940
          G  R  T  M  H  D  L  M  A  L  D  E  A  L  R  L  P  N  H
L
                .        .        .        .        .
.
   5941
GCCGTCCTGCAACCGGTAGACCGGCTGGGCCTACACGGCCCAGACGTACGGCTCCATCTC   6000
          P  L  V  N  A  M  Q  G  V  R  I  H  R  T  Q  M  G  L  Y
L
                .        .        .        .        .
.
   6001
GGGGTCGTACTAGCCCAACAACCTCTGGAGCTTTGGGAGCCACACCTTCACCACGAGCCA   6060
          G  L  M  I  P  N  N  S  V  E  F  G  E  T  H  F  H  H  E
T
                .        .        .        .        .
.
   6061
CTTCCTGTCAGGGGTCATCGGCTCAAGCAGCCGGCGGACGCGGACGGCCCACTCGACGGC   6120
          F  S  L  G  W  Y  G  L  E  D  A  A  Q  A  Q  R  T  L  Q
R
                .        .        .        .        .
.
   6121
CTCGTACAAGACCATCAAGACGCCTAACTGGGGGCGGTATGGGGCGACCTGGACGCGTAC   6180
          L  M  N  Q  Y  N  Q  P  N  V  G  A  M  G  R  Q  V  Q  A
H
                .        .        .        <orfdwn2 .
.
   6181
ACTGCCGACCGTTGGCAGATAGAAGAGAATGGACTTCACCCTGGCTCCTCCGGTTCGCGG   6240

TGACGGCTGGCAACCGTCTATCTTCTCTTACCTGAAGTGGGACCGAGGAGGCCAAGCGCC
          S  G  V  T  P  L  Y  F  L  I  S  K  M
                .        .        .        .        .
.
```

```
   6241
CGCCCTCCATTGACGTGCGCCGAAAGCGGCTCGACCGTCCCACTCCGCCCTTGAGTTCCG   6300

GCGGGAGGTAACTGCACGCGGCTTTCGCCGAGCTGGCAGGGTGAGGCGGGAACTCAAGGC
5                  .         .         .         .         .
.
   6301
TCTGACGCCGCGCCAGTCGGCGGGCCGTCCGCCGGGGTGCCCGCCGGGGTCCGCACCCGC   6360

10 AGACTGCGGCGCGGTCAGCCGCCCGGCAGGCGGCCCCACGGGCGGCCCCAGGCGTGGGCG
                   .         .         .         .         .
.
   6361
CGGACGGCACGGCGCGCACCGCGCGCGCGGCGCTTCGGGGCACCGGGCTCGACGGGGTGC   6420
15
GCCTGCCGTGCCGCGCGTGGCGCGCGCGCCGCGAAGCCCCGTGGCCCGAGCTGCCCCACG
                   .         .         .         .         .
.
   6421
20 TCAGCGGGACGTCCAACGGAAGGCAAGCCCCCGTACCCAGCCTGGTCAAGGCGCTCATCG   6480

AGTCGCCCTGCAGGTTGCCTTCCGTTCGGGGGCATGGGTCGGACCAGTTCCGCGAGTAGC

                                                    orfdwn3>
25                 .         .         .         .         .  M  P
G
   6481
CCATTCCCTGAGGAGGTCCCGCCTTGACCACAGCAATCTCCGCGCTCCCGACCGTGCCCG   6540

30 GGTAAGGGACTCCTCCAGGGCGGAACTGGTGTCGTTAGAGGCGCGAGGGCTGGCACGGGC
                   .         .         .         .         .
.
   6541
GCTCCGGACTCGAAGCACTGGACCGTGCCACCCTCATCCACCCCACCCTCTCCGGAAACA   6600
35            S  G  L  E  A  L  D  R  A  T  L  I  H  P  T  L  S  G  N
T
                   .         .         .         .         .
.
   6601
40 CCGCGGAACGGATCGTGCTGACCTCGGGGTCCGGCAGCCGGGTCCGCGACACCGACGGCC   6660
            A  E  R  I  V  L  T  S  G  S  G  S  R  V  R  D  T  D  G
R
                   .         .         .         .         .
.
```

```
   6661
GGGAGTACCTGGACGCGAGCGCCGTCCTCGGGGTGACCCAGGTGGGCCACGGCCGGGCCG   6720
          E  Y  L  D  A  S  A  V  L  G  V  T  Q  V  G  H  G  R  A
E
5                 .        .        .        .        .
.
   6721
AGCTGGCCCGGGTCGCGGCCGAGCAGATGGCCCGGCTGGAGTACTTCCACACCTGGGGGA   6780
          L  A  R  V  A  A  E  Q  M  A  R  L  E  Y  F  H  T  W  G
10 T
                  .        .        .        .        .
.
   6781
CGATCAGCAACGACCGGGCGGTGGAGCTGGCGGCACGGCTGGTGGGGCTGAGCCCGGAGC   6840
15        I  S  N  D  R  A  V  E  L  A  A  R  L  V  G  L  S  P  E
P
                  .        .        .        .        .
.
   6841
20 CGCTGACCCGCGTCTACTTCACCAGCGGCGGGGCCGAGGGCAACGAGATCGCCCTGCGGA   6900
          L  T  R  V  Y  F  T  S  G  G  A  E  G  N  E  I  A  L  R
M
                  .        .        .        .        .
.
25    6901
TGGCCCGGCTCTACCACCACCGGCGCGGGGAGTCCGCCCGTACCTGGATACTCTCCCGCC   6960
          A  R  L  Y  H  H  R  R  G  E  S  A  R  T  W  I  L  S  R
R
                  .        .        .        .        .
30 .
   6961
GGTCGGCCTACCACGGCGTCGGATACGGCAGCGGCGGCGTCACCGGCTTCCCCGCCTACC   7020
          S  A  Y  H  G  V  G  Y  G  S  G  G  V  T  G  F  P  A  Y
H
35                .        .        .        .        .
.
   7021
ACCAGGGCTTCGGCCCCTCCCTCCCGGACGTCGACTTCCTGACCCCGCCGCAGCCCTACC   7080
          Q  G  F  G  P  S  L  P  D  V  D  F  L  T  P  P  Q  P  Y
40 R
                  .        .        .        .        .
.
   7081
GCCGGGAGCTGTTCGCCGGTTCCGACGTCACCGACTTCTGCCTCGCCGAACTGCGCGAGA   7140
```

R E L F A G S D V T D F C L A E L R E

T

Sau

5 7141 CCATCGACCGGATCGGCCCGGAGCGGATCGCGGCGATGATCGGCGAGCCGATC

I D R I G P E R I A A M I G E P I